An Empirical Study on Adaptive Learning Using Decision Trees for Student Performance Prediction

Asia Bataineh
Computer Science Department
Faculty of Information Technology
University of Petra (UoP)
Amman, Jordan
asia.bataineh@uop.edu.jo

Ahlam Al-E'mari
Department of Finance
School of Business
University of Jordan (UJ)
Amman, Jordan
ahlam.r.ammari@gmail.com

Rula Hajjaj
Department of Business Administration
The Applied College
Northern border University
Arar, Saudi Arabia
2370257301@nbu.edu.sa

Hazem Bani Abdoh
Computer Science Department
Faculty of Computer Science and Information Technology
Jerash university
Jerash, Jordan
Hazim.baniabdoh@jpu.edu.jo

Fuad Fataftah
Mutlimedia systems and Virtual Reality
School of Computer science
Universiti Sains Malaysia (USM)
Penang, Malaysia
fuadmanna@student.usm.my

***Abstract*—This study investigates the application of adaptive decision tree models for predicting student performance, focusing on their role in educational data mining and adaptive learning. A major challenge in student assessment is identifying at-risk learners and dynamically adjusting predictive models to accommodate evolving learning patterns. Traditional static models often fail to capture these variations, highlighting the need for adaptive approaches. This research develops an adaptive decision tree framework that integrates incremental learning and adaptive parameter tuning to improve predictive accuracy and stability. Using the Student Performance dataset from the UCI Machine Learning Repository, the study applies multiple decision tree variants, including C4.5, CART, Random Forest, and Gradient Boosted Trees. The dataset, consisting of demographic, social, and academic attributes of students, is preprocessed and split into 70% training and 30% testing sets. Models are evaluated using accuracy, precision, recall, and F1-score to assess their effectiveness. The results demonstrate that adaptive decision tree models significantly outperform static models, with the best adaptive model (Gradient Boosted Decision Trees) achieving an accuracy of 84% and an F1-score of 82%, compared to 77% and 74%, respectively, for static counterparts. These findings highlight the potential of adaptive learning models to enhance personalized learning interventions and support data-driven curriculum design. This paper concludes by discussing the implications of adaptive decision tree applications in real-world educational settings and proposes future directions for improving predictive modeling in adaptive learning environments.**

***Index Terms*—Adaptive Learning, Student Performance Prediction, Decision Trees, Machine Learning, Educational Data Mining.**

I. Introduction

The increasing availability of digital educational data and the widespread adoption of learning management systems have led to significant growth in Educational Data Mining (EDM), enabling institutions to derive insights into student learning patterns and academic performance [1]. Predicting student success is a crucial aspect of EDM, as it allows educators to implement timely interventions, improve personalized learning, and enhance curriculum design [2]. However, despite the advancements in machine learning, accurately forecasting student outcomes remains challenging due to temporal shifts in learning behaviors, contextual variations, and the evolving nature of academic performance.

Decision tree-based machine learning models have been widely adopted for classification and prediction tasks in education due to their interpretability, robustness, and computational efficiency [3]. Traditional models such as C4.5 and CART have been used to identify at-risk students, predict test performance, and assess dropout risks. More recent methods, including Random Forest and Gradient Boosted Trees, have further improved predictive accuracy by leveraging ensemble learning. However, a major limitation of these approaches is their static nature—once trained, they rely on predefined features and parameters that may not adapt to evolving educational data. Student performance is dynamic, influenced by factors such as curriculum changes, pedagogical interventions, external circumstances, and student engagement levels. Traditional models fail to capture these variations over time, leading to performance degradation and reduced reliability in long-term applications [4].

State-of-the-art research in EDM has begun to explore adaptive learning techniques to overcome these challenges. Incremental learning methods, which allow models to update dynamically as new data becomes available, have shown promise in handling concept drift—a phenomenon where the underlying data distribution changes over time. However, existing adaptive approaches primarily focus on deep learning models, which, despite their high accuracy, often suffer from a lack of interpretability and require large datasets. This study addresses this gap by integrating adaptive learning mechanisms into decision tree models, enabling them to adjust dynamically to evolving student performance patterns while maintaining model transparency and efficiency [5].

This research proposes an adaptive decision tree framework that incorporates incremental learning and adaptive parameter tuning to improve predictive accuracy and model stability in educational settings. By evaluating multiple decision tree variants (C4.5, CART, Random Forest, and Gradient Boosted Trees) and implementing adaptive mechanisms, this study aims to bridge the gap between traditional static models and highly complex deep learning systems. The findings of this research will contribute to personalized learning strategies, data-driven decision-making in education, and the development of predictive models that remain effective over time.

The remainder of this paper is organized as follows: Section II reviews related work in educational data mining and decision tree modeling. Section III details the dataset, preprocessing, and feature engineering steps. Section IV describes the proposed adaptive model framework, including the algorithms and adaptive mechanisms employed. Section IV presents experimental results, while Section V discusses the findings and limitations. Section VI concludes the paper and outlines future research directions.

II. Literature Review

Educational data mining (EDM) has evolved significantly, leveraging machine learning techniques to predict student performance and enhance personalized learning strategies. Previous studies have primarily focused on feature-based modeling, dropout prediction, and performance classification using machine learning approaches such as logistic regression, support vector machines (SVMs), and deep learning techniques [6]–[8]. While these methods have achieved notable success, they often present trade-offs in terms of interpretability, computational efficiency, and adaptability to changing learning environments. Predicting academic performance often involves modeling student test scores, grades, or dropout probabilities [9]. Earlier research adopted linear regression and logistic regression models [10], while more recent trends incorporate machine learning classifiers, including Support Vector Machines, Neural Networks, and Ensemble methods [11].In addition, features used for predicting performance span a broad range, from demographic variables and parental education levels to attendance records, study habits, and social interactions [12].

Decision tree models have been widely used in EDM due to their transparency, ease of implementation, and ability to handle categorical and numerical data. Studies utilizing traditional decision tree approaches, including C4.5 and CART, have demonstrated their effectiveness in predicting student success based on demographic, academic, and behavioral attributes, and the UCI Student Performance dataset has been utilized as a benchmark to evaluate the efficacy of different classifications algorithms [13], [14]. However, these models rely on static training data, which limits their ability to accommodate temporal changes in student learning behaviors. Ensemble methods, such as Random Forest and Gradient Boosted Trees (GBDT), have been introduced to improve accuracy by reducing variance and overfitting, yet they still suffer from fixed parameter configurations and do not inherently adapt to evolving educational datasets [15], [16].

To address the limitations of static models, adaptive learning approaches have been proposed, particularly in the context of deep learning and online learning frameworks. Incremental learning, which allows models to update dynamically as new data becomes available, has been shown to improve predictive accuracy in rapidly changing environments [17], [18]. However, most adaptive models in EDM focus on deep neural networks (DNNs), recurrent neural networks (RNNs), and transformers, which, despite their strong predictive capabilities, lack interpretability and require extensive computational resources [19], [20].

Concept drift, a major challenge in educational datasets, further complicates student performance prediction. Learning patterns evolve over time due to curriculum changes, varying instructional methods, and shifting student engagement. Some studies have proposed adaptive approaches to mitigate this issue, such as online learning algorithms and reinforcement learning models [21]. However, these methods often demand large-scale datasets and continuous model retraining, making them impractical for real-time educational applications.

Despite the growing emphasis on adaptive learning in EDM, few studies have systematically explored adaptive decision tree models as a middle-ground solution. Adaptive decision trees can incorporate incremental learning mechanisms while maintaining transparency and computational efficiency, making them a viable alternative to deep learning-based models. This research builds on prior work by introducing an adaptive decision tree framework that dynamically adjusts parameters and integrates incremental updates to handle temporal changes in student performance data. By comparing multiple decision tree variants with and without adaptation, this study aims to demonstrate the advantages of adaptive decision trees over traditional static models and position them as a practical tool for real-time student performance prediction.

III. Methodology

Our proposed framework integrates adaptive learning mechanisms into decision tree models to enhance their predictive accuracy and robustness. The framework comprises four main components: Data Preprocessing, Base Decision Tree Model, Adaptive Parameter Tuning, and Incremental Learning. Figure 1 presents a block diagram of the proposed system.

A. Data Preparation

1) Dataset Description: The Student Performance dataset used in this study is sourced from the UCI Machine Learning Repository [5]. It includes data from Portuguese secondary school students, comprising features related to demographic, social, and academic attributes. The dataset covers two subjects—Mathematics and Portuguese—collected from two dif-

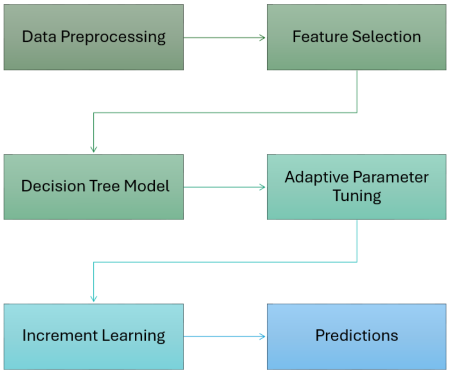

Fig. 1. Block Diagram of the Proposed Adaptive Decision Tree Framework

ferent schools. For this study, we focus on the Mathematics dataset, which includes 395 students. Moreover, the dataset consists of 33 features, including:

- Demographic and Social Attributes: Gender, age, parent's education, family relationships, and travel time.
- Academic Attributes: Study time, past exam scores, failures in previous classes, and weekly alcohol consumption.
- School-Related Factors: Extra educational support, internet access, and attendance to extra paid classes.

The target variable is the final grade (G3), which is a numeric grade between 0 and 20. For classification purposes, following prior studies [14], we discretized the G3 score into three categories: Low (0–9), Medium (10–14), and High (15–20).

2) Data Preprocessing:

- Missing Values: The dataset is relatively complete, but any rows with missing values were dropped, resulting in a final dataset size of 395 complete records.
- Categorical Encoding: Some features, such as school (binary category) and sex (binary category), were one-hot encoded. Multi-level categorical features (e.g., parent's job, internet access) were also converted into dummy variables.
- Feature Scaling: Although decision trees are less sensitive to scaling, we standardized numeric features to facilitate comparability and improve the stability of incremental learning steps.
- Class distribution across the three categories (Low, Medium, High) was slightly imbalanced, with fewer students in the High category. To address this, we applied SMOTE (Synthetic Minority Over-sampling Technique) on the training set to balance the classes [22]. This step ensures that the decision tree models do not become biased toward the majority class.
- Splitting Data: The data was split into training (70%) and testing (30%) sets. Additionally, a validation set (10% of training data) was extracted for adaptive parameter tuning. A temporal splitting approach was not strictly necessary as the dataset does not inherently represent a time series, but we simulate incremental learning scenarios by feeding data gradually during model updates.

3) Feature Selection: We applied a mutual information-based feature selection to reduce the dimensionality and identify the most influential features [23]. Features like past academic performance (G1, G2), study time, failures, and parent's education level emerged as top predictors. A reduced feature set of 20 attributes was selected to minimize model complexity without sacrificing predictive performance.

TABLE I
SELECTED FEATURES

Feature	Type	Description
G1 (Past Grade)	Numeric	First period grade
G2 (Past Grade)	Numeric	Second period grade
studytime	Numeric	Weekly study time
failures	Numeric	Number of past class failures
famrel (1-5)	Numeric	Quality of family relationships
Medu	Numeric	Mother's education level (0-4)
Fedu	Numeric	Father's education level (0-4)
internet	Categorical	Internet access at home
absences	Numeric	Number of school absences
schoolsup	Categorical	Extra educational support (yes/no)
activities	Categorical	Extra-curricular activities (yes/no)
higher	Categorical	Wants to take higher education (yes/no)
romantic	Categorical	With a romantic relationship (yes/no)
Walc	Numeric	Weekend alcohol consumption (1-5)
Dalc	Numeric	Workday alcohol consumption (1-5)
Pstatus	Categorical	Parent cohabitation status (A,T)
traveltime	Numeric	Home to school travel time (1-4)
age	Numeric	Age of the student
sex	Categorical	Gender of the student (F/M)
paid	Categorical	Extra paid classes (yes/no)

B. Decision Tree Models

We evaluate multiple decision tree variants as our base model:

- C4.5 Decision Tree: A well-known algorithm that uses information gain ratio for splitting and handles both numeric and categorical features [24].
- Classification and Regression Trees (CART): A widely used algorithm that employs Gini impurity or entropy for node splitting [25].
- Ensemble Methods: We also consider Random Forest [26] and Gradient Boosted Decision Trees (GBDT) [27]. These methods combine multiple trees to reduce variance and potentially improve accuracy.

C. Adaptive Parameter Tuning

Adaptive parameter tuning involves dynamically adjusting hyperparameters—such as tree depth, minimum samples per leaf, and learning rate (for ensembles)—in response to model performance on a validation set. We implement a feedback loop that regularly evaluates the model on a small, held-out validation subset. If the performance declines or stabilizes below a threshold, the tuner adjusts parameters to explore different configurations [28]. This process continues until optimal or stable performance is achieved.

D. Incremental Learning Approach

To simulate real-world scenarios where data arrive continuously over time, we employ incremental learning. At scheduled intervals (e.g., after every batch of new student data or at the end of an academic quarter), the model updates its parameters with the newly available data. Incremental learning involves partial fitting of the model, enabling it to incorporate new patterns without forgetting previously learned information [29]. This approach is particularly relevant for educational contexts, where student behavior and learning environments evolve.

E. Evaluation Metrics

We use accuracy, precision, recall, and F1-score to evaluate our classification models. To provide more nuanced insights, we also consider a weighted F1-score to handle class imbalance. Furthermore, we conduct a feature importance analysis using metrics such as Gini importance or Shapley values to determine which features most strongly influence the predictions [30].

IV. Experimental Setup and Results

The experiments were conducted using Python's scikit-learn library on a workstation equipped with an Intel Core i7 processor and 16GB of RAM. A 10-fold cross-validation strategy was employed to ensure robustness and minimize the risk of bias from a single train-test split. Each model was trained and tested across multiple folds, with performance metrics averaged to provide reliable evaluations. To evaluate the adaptive component, incremental learning was simulated by sequentially feeding batches of 50 students, emulating real-world data availability. After each batch, model parameters were tuned, and partial fitting was performed, with performance metrics recorded before and after adaptation.

We computed feature importance using Gini importance from the best-performing model (Adaptive GBDT). Figure 2 lists the top 10 features and their relative importance scores. Where the past academic performance (G2, G1) and studytime are the top predictors, consistent with educational intuition. Parental education levels and absences also significantly influence final grades

The baseline models included both traditional decision tree algorithms and ensemble methods. Static C4.5 and CART decision trees were used without adaptive parameter tuning or incremental learning mechanisms. Ensemble models, such as Random Forest and Gradient Boosting Decision Trees (GBDT), were evaluated using default parameters and trained on the initial dataset without further adaptation. Furthermore, adaptive mechanisms were introduced to each base model to enhance their predictive capabilities. As baselines, we included:

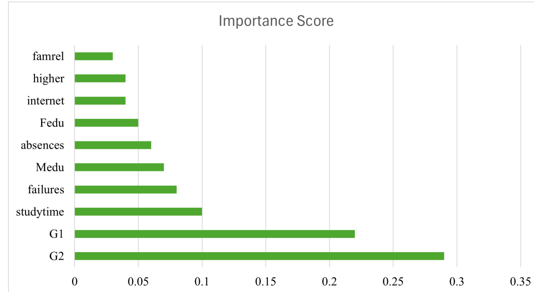

Fig. 2. Top 10 Features by importance.

- Static C4.5 Decision Tree: No adaptive parameter tuning or incremental updates.
- Static CART Decision Tree: Similarly, no adaptive mechanisms.
- Ensemble Models (Random Forest, GBDT) without Adaptation: Using default parameters and training on the initial dataset only.

The Adaptive C4.5 and CART models incorporated incremental learning and parameter tuning. Similarly, the Adaptive Random Forest and Adaptive GBDT models featured adaptive parameter adjustments and updates to accommodate incremental data. These enhancements aimed to improve model accuracy and generalization under dynamic data conditions.

The comparative performance metrics for baseline and adaptive models are summarized in Table II. The results reveal a clear advantage for adaptive models over their static counterparts. While the static GBDT achieved the best performance among non-adaptive models (accuracy: 0.77, F1-score: 0.74), the Adaptive GBDT surpassed all other models with an accuracy of 0.84 and an F1-score of 0.82. Similarly, Adaptive Random Forest and Adaptive C4.5 demonstrated significant improvements in precision, recall, and F1-score compared to their static versions.

TABLE II
Performance Metrics of Baseline vs. Adaptive Models

Model	Accuracy	Precision	Recall	F1-score
Static C4.5	0.72	0.70	0.68	0.69
Static CART	0.70	0.68	0.65	0.66
Static Random Forest	0.75	0.73	0.71	0.72
Static GBDT	0.77	0.75	0.72	0.74
Adaptive C4.5	0.79	0.77	0.76	0.76
Adaptive CART	0.78	0.74	0.75	0.74
Adaptive Random Forest	0.82	0.80	0.79	0.79
Adaptive GBDT	0.84	0.82	0.81	0.82

These results highlight the benefits of incorporating adaptive learning techniques, including incremental updates and parameter tuning. Adaptive models not only achieved higher accuracy but also demonstrated better generalization, making them well-suited for dynamic and evolving educational datasets. The findings underscore the potential of adaptive mechanisms in enhancing the performance and applicability of machine learning models in real-world educational settings.

V. Discussion

Decision trees are inherently interpretable, allowing educational stakeholders to understand how features affect student performance. The adaptive approach enhances predictive accuracy without compromising interpretability. For instance, an educator can see that poor study habits (low studytime) and multiple past failures are strong indicators of low final grades, prompting targeted interventions.

The adaptive methodology addresses the concept drift that can occur over the academic year. Changes in the curriculum, teaching strategies, or student demographics may alter performance patterns. By continuously updating the model as new student data arrive, the predictive system remains relevant and accurate. This adaptability is especially useful in online learning platforms and MOOCs, where cohorts may differ significantly from one session to another.

Previous studies using static models on the Student Performance dataset reported accuracies ranging from 0.65 to 0.78 for various methods [14]. Our adaptive GBDT model achieving 0.84 accuracy indicates a notable improvement. Integrating adaptive tuning aligns with recent trends in machine learning research, where incremental and online learning methods are sought to maintain model relevance over time [18], [20].

While the results are promising, some limitations merit consideration:

1) Data Size: The Student Performance dataset is relatively small and may not capture large-scale educational complexities.
2) Temporal Simulation: We simulated incremental learning by feeding batches of data. Real-time adaptation would require a streaming environment.
3) Generality: The model's performance and adaptive capabilities may differ when applied to other datasets, subjects, or educational contexts.

While the results are promising, some limitations merit consideration. One of the key challenges faced in this study was the relatively small dataset size, which may not fully capture the complexities of large-scale educational environments. Additionally, implementing incremental learning in a real-time educational setting remains a challenge, as it requires continuous data availability and robust adaptation mechanisms. Future research should explore larger datasets and real-time applications to further validate the effectiveness of adaptive decision tree models.

In addition, decision tree models have certain limitations compared to neural network-based approaches. Deep learning models, such as artificial neural networks (ANNs), recurrent neural networks (RNNs), and transformer-based models, have been widely used for student performance prediction due to their ability to model complex, nonlinear relationships within educational data [31]. RNNs and Long Short-Term Memory (LSTM) networks, for example, can effectively capture sequential patterns in student learning behaviors and adapt dynamically based on historical data [32]. Transformer-based architectures, such as BERT and its educational adaptations, have demonstrated strong performance in text-based learning analytics and personalized recommendation systems [33].However, these neural network-based models require significantly larger datasets, extensive hyperparameter tuning, and high computational resources, making them less feasible for real-time, low-resource educational environments [34].

Future research could explore hybrid models that integrate decision trees with neural network architectures, such as combining decision trees with deep learning frameworks through hybrid ensemble learning or embedding-based feature engineering, to leverage the strengths of both approaches. This could help improve predictive accuracy while maintaining a degree of interpretability for practical educational applications.

VI. Conclusion

This study proposed an adaptive decision tree framework for student performance prediction, integrating incremental learning and adaptive parameter tuning to enhance predictive accuracy. The findings demonstrated that adaptive models significantly outperformed static decision trees, with the Adaptive Gradient Boosted Decision Trees achieving the highest accuracy (84%) and F1-score (82%). These results underscore the importance of dynamic modeling approaches in educational data mining, as they allow for real-time adjustments to shifting student learning behaviors. The study's implications extend beyond predictive analytics, providing educators with actionable insights for personalized learning interventions and curriculum optimization. However, the research faced limitations, including the relatively small dataset and the simulated nature of incremental learning, which may not fully capture real-time educational dynamics. Future research should focus on applying this framework to larger and more diverse datasets, implementing real-time adaptation mechanisms in live educational environments, and exploring hybrid models that combine decision trees with deep learning techniques to balance interpretability and predictive power. These advancements would further improve the robustness and practical applicability of adaptive learning models in education.

References

[1] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews)*, vol. 40, no. 6, pp. 601–618, 2010.

[2] R. Baker, "Data mining for education," *International Encyclopedia of Education/Elsevier*, 2010.

[3] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[4] Z. Papamitsiou and A. A. Economides, "Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence," *Journal of Educational Technology & Society*, vol. 17, no. 4, pp. 49–64, 2014.

[5] D. Dua and C. Graff, "Uci machine learning repository-student performance," 2017.

[6] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. Baker, *Handbook of educational data mining*. CRC press, 2010.

[7] B. E. Schools, "Using student data to improve teaching and learning," *BE Schools*, 2008.

[8] J. Vulperhorst, C. Lutz, R. de Kleijn, and J. van Tartwijk, "Disentangling the predictive validity of high school grades for academic success in university," *Assessment & Evaluation in Higher Education*, vol. 43, no. 3, pp. 399–414, 2018.

[9] S. Kulkarni and G. Harman, *An elementary introduction to statistical learning theory*. John Wiley & Sons, 2011.

[10] D. A. Shafiq, M. Marjani, R. A. A. Habeeb, and D. Asirvatham, "Student retention using educational data mining and predictive analytics: a systematic literature review," *IEEE Access*, vol. 10, pp. 72 480–72 503, 2022.

[11] P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," *dsi.uminho.pt*, 2008.

[12] A. Gelman, *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2007.

[13] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Predicting students'performance in distance learning using machine learning techniques," *Applied Artificial Intelligence*, vol. 18, no. 5, pp. 411–426, 2004.

[14] Z. Papamitsiou and A. A. Economides, "Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence," *Journal of Educational Technology & Society*, vol. 17, no. 4, pp. 49–64, 2014.

[15] G. W. Dekker, M. Pechenizkiy, and J. M. Vleeshouwers, "Predicting students drop out: A case study," in *Proceedings of the 2nd International Conference on Educational Data Mining, EDM 2009, July 1-3, 2009. Cordoba, Spain*, 2009, pp. 41–50.

[16] A. Van den Beemt, J. Buijs, and W. Van der Aalst, "Analysing structured learning behaviour in massive open online courses (moocs): an approach based on process mining and clustering," *International Review of Research in Open and Distributed Learning*, vol. 19, no. 5, 2018.

[17] A. Tüfekci and E. A. G. Yılmaz, "Educational data mining: A systematic literature mapping study," *Engineering Education Trends in the Digital Era*, pp. 70–82, 2020.

[18] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM computing surveys (CSUR)*, vol. 46, no. 4, pp. 1–37, 2014.

[19] G. Widmer and M. Kubat, "Learning in the presence of concept drift and hidden contexts," *Machine learning*, vol. 23, pp. 69–101, 1996.

[20] Y. Zhang, P. Zhao, S. Niu, Q. Wu, J. Cao, J. Huang, and M. Tan, "Online adaptive asymmetric active learning with limited budgets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 6, pp. 2680–2692, 2019.

[21] B. Sonnleitner, T. Madou, M. Deceuninck, F. Theodosiou, and Y. R. Sagaert, "Evaluation of early student performance prediction given concept drift," *Computers and Education: Artificial Intelligence*, p. 100369, 2025.

[22] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[23] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on neural networks*, vol. 5, no. 4, pp. 537–550, 1994.

[24] Z. Huang, T. D. Gedeon, and M. Nikravesh, "Pattern trees induction: A new machine learning method," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 4, pp. 958–970, 2008.

[25] L. Wilkinson, "Classification and regression trees," *Systat*, vol. 11, pp. 35–56, 2004.

[26] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random forests," *Ensemble machine learning: Methods and applications*, pp. 157–175, 2012.

[27] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[28] R. Bardenet, M. Brendel, B. Kégl, and M. Sebag, "Collaborative hyperparameter tuning," in *International conference on machine learning*. PMLR, 2013, pp. 199–207.

[29] A. Bifet and R. Gavalda, "Adaptive learning from evolving data streams," in *Advances in Intelligent Data Analysis VIII: 8th International Symposium on Intelligent Data Analysis, IDA 2009, Lyon, France, August 31-September 2, 2009. Proceedings 8*. Springer, 2009, pp. 249–260.

[30] S. Lundberg, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.

[31] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Pearson, 2016.

[32] B. Alnasyan, M. Basheri, and M. Alassafi, "The power of deep learning techniques for predicting student performance in virtual learning environments: A systematic literature review," *Computers and Education: Artificial Intelligence*, p. 100231, 2024.

[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need. advances in neural information processing systems," *Advances in neural information processing systems*, vol. 30, no. 2017, 2017.

[34] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.